Leveraging molecular structure and bioactivity with chemical language models for de novo drug design

Year: 2023 | Citations: 120 | Authors: Michael Moret, Irene Pachon Angona, Leandro Cotos, Shen Yan, Kenneth Atz

Abstract

Generative Deep Learning holds promise for mining the unexplored “chemical universe” for new drugs. Here, the authors demonstrate the de novo design of phosphoinositide 3-kinase gamma (PI3Kγ) inhibitors for the PI3K/Akt pathway in human tumor cells. Generative chemical language models (CLMs) can be used for de novo molecular structure generation by learning from a textual representation of molecules. Here, we show that hybrid CLMs can additionally leverage the bioactivity information available for the training compounds. To computationally design ligands of phosphoinositide 3-kinase gamma (PI3Kγ), a collection of virtual molecules was created with a generative CLM. This virtual compound library was refined using a CLM-based classifier for bioactivity prediction. This second hybrid CLM was pretrained with patented molecular structures and fine-tuned with known PI3Kγ ligands. Several of the computer-generated molecular designs were commercially available, enabling fast prescreening and preliminary experimental validation. A new PI3Kγ ligand with sub-micromolar activity was identified, highlighting the method’s scaffold-hopping potential. Chemical synthesis and biochemical testing of two of the top-ranked de novo designed molecules and their derivatives corroborated the model’s ability to generate PI3Kγ ligands with medium to low nanomolar activity for hit-to-lead expansion. The most potent compounds led to pronounced inhibition of PI3K-dependent Akt phosphorylation in a medulloblastoma cell model, demonstrating efficacy of PI3Kγ ligands in PI3K/Akt pathway repression in human tumor cells. The results positively advocate hybrid CLMs for virtual compound screening and activity-focused molecular design.